

These amendments are for discussion purposes only. /KRS/

**From: John Garrity**

Ref. 10/797,359 - our telephonic Interview today at 11am - I will call you

Topic 1 - I would like to review the attached claim amendments to get your opinion if they address the 35 USC 112 (first and second paragraph) and the 35 USC 101 rejections.

Topic 2 – possible amendments to further prosecution towards allowance – more specifically details in claims related to steps A – G on pages 15-16 of the application.

Thank you very much for your time,

**John Garrity**  
(203) 925-9400 x39

**DO NOT ENTER /KRS/ 06/25/2009**

These amendments are for discussion purposes only. /KRS/

1. (Previously Presented) A method to process a text document, comprising:  
partitioning text of the text document and assigning semantic meaning to words of the partitioned text, where assigning comprises applying a plurality of regular expressions, rules and a plurality of dictionaries to recognize chemical name fragments;  
  
recognizing any substructures present in the chemical name fragments;  
  
determining structural connectivity information of the chemical name fragments and recognized substructures;  
  
extracting information associated with the recognized chemical name fragments and substructures of the text document and indexing the extracted information in a text index;  
  
indexing representations of the recognized chemical name fragments and the substructures in association with the determined structural connectivity information into a plurality of chemical connectivity tables;  
  
storing the text index in association with the indexed representations in a searchable index; and  
  
providing a graphical user interface to search the searchable index, where the search comprises entering one or more chemical fragment names and entering one or more substructures in a representation form, where the entering is by at least one of text form or graphical selection.

2. (Currently Amended) A The method as in claim 1, wherein the extracting further comprises extracting keywords from the text document and indexing the keywords in the text index, and

These amendments are for discussion purposes only. /KRS/

These amendments are for discussion purposes only. /KRS/

wherein the search comprises selecting a graphical representation of one or more substructures and additionally entering at least one keyword.

3. (Currently Amended) ~~A~~ The method as in claim 1, wherein extracting further comprises extracting keywords from the text document and indexing the keywords in the text index, and wherein the search comprises additionally entering at least one keyword, and at least one of chemical name fragment connectivity and substructure connectivity.

4. (Canceled)

5. (Canceled)

6. (Currently Amended) ~~A~~ The method as in claim 1, wherein the search further comprises entering at least one search term, and where a search results in an intersection of the indexed representations and the text index, identifying at least one document that contains a reference to a corresponding chemical compound.

7. (Currently Amended) ~~A~~ The method as in claim 1, where determining structural connectivity information comprises looking up recognized chemical name fragments and substructures in a structure dictionary.

8. (Currently Amended) ~~A~~ The method as in claim 1, where the representations comprise MOL type representations and SMILES type representations.

9. (Currently Amended) ~~A~~ The method as in claim 1, where said plurality of dictionaries

These amendments are for discussion purposes only. /KRS/

comprise a dictionary of common chemical prefixes and a dictionary of common chemical suffixes.

10. (Currently Amended) A The method as in claim 1, where said plurality of dictionaries comprise a dictionary of stop words to eliminate erroneous chemical name fragments.

11. (Currently Amended) A The method as in claim 1, further comprising filtering recognized chemical name fragments using a list of stop words to eliminate erroneous chemical name fragments.

12. (Currently Amended) A The method as in claim 1, where chemical name fragments are further recognized by using common chemical word endings.

13. (Currently Amended) A The method as in claim 1, where application of said regular expressions and rules results in punctuation characters being one of maintained or removed between chemical name fragments as a function of context.

14. (Currently Amended) A The method as in claim 1, where said regular expressions comprise a plurality of patterns, individual ones of which are comprised of at least one of characters, numbers and punctuation.

15. (Currently Amended) A The method as in claim 14, where the punctuation comprises at least one of parenthesis, square bracket, hyphen, colon and semi-colon.

16. (Currently Amended) A The method as in claim 14, where the characters comprise ~~at least~~

These amendments are for discussion purposes only. /KRS/

~~one of upper case C, O, R, N and H.~~

17. (Currently Amended) ~~A~~ The method as in claim 14, where the characters comprise strings of at least one of lower case xy, ene, ine, yl, ane and oic.

18. (Currently Amended) ~~A~~ The method as in claim 1, comprising an initial step of tokenizing the document to provide a sequence of tokens.

19. (Currently Amended) ~~A~~ system to process a text document An apparatus, comprising:

~~a unit tokenizer module and a token processing module configured to partition text of the a text document and to assign semantic meaning to words of the partitioned text, where assigning comprises applying a plurality of regular expressions, rules and a plurality of dictionaries to recognize chemical name fragments;~~

~~a unit the token processing module configured to recognize any substructures present in the chemical name fragments;~~

~~a unit the token processing module configured to extract information associated with the recognized chemical name fragments and substructures of the text document and index the extracted information in a text index;~~

~~a unit the token processing module configured to determine structural connectivity information of the chemical name fragments and recognized substructures and to index representations of the chemical name fragments and the recognized substructures in association with the determined~~

These amendments are for discussion purposes only. /KRS/

structural connectivity information into a plurality of chemical connectivity tables;

~~a unit to store the text index in association with indexed representations in a searchable index; and~~

~~a unit to provide a searcher module and a graphical user interface configured to search the searchable index, where the search comprises entering one or more chemical fragment names and entering one or more substructures in a representation form, where the entering is by at least one of text form or graphical selection.~~

20. (Currently Amended) ~~A system~~ The apparatus as in claim 19, wherein the ~~unit to extract token processing module~~ is further configured to extract keywords from the text document and index the keywords in the text index, and wherein the search comprises selecting a graphical representation of one or more substructures and additionally entering at least one keyword.

21. (Currently Amended) ~~A system~~ The apparatus as in claim 19, wherein the ~~unit to extract token processing module~~ is further configured to extract keywords from the text document and index the keywords in the text index, and wherein the search comprises additionally entering at least one keyword, and at least one of chemical name fragment connectivity and substructure connectivity.

22. (Canceled)

23. (Canceled)

These amendments are for discussion purposes only. /KRS/

24. (Currently Amended) ~~A system~~ The apparatus as in claim 19, wherein the search further comprises entering at least one search term and wherein a search results in an intersection of the indexed representations and the text index, to identify at least one document that contains a reference to a corresponding chemical compound.
25. (Currently Amended) ~~A system~~ The apparatus as in claim 19, where said ~~unit that determines structural connectivity information token processing module is further configured to~~ looks up recognized fragments and substructures in a structure dictionary.
26. (Currently Amended) ~~A system~~ The apparatus as in claim 19, where the representations comprise MOL type representations and SMILES type representations.
27. (Currently Amended) ~~A system~~ The apparatus as in claim 19, where said plurality of dictionaries comprise a dictionary of common chemical prefixes and a dictionary of common chemical suffixes.
28. (Currently Amended) ~~A system~~ The apparatus as in claim 19, where said plurality of dictionaries comprise a dictionary of stop words to eliminate erroneous chemical name fragments.
29. (Currently Amended) ~~A system~~ The apparatus as in claim 19, further comprising a ~~unit~~ said token processing module is further configured to filter recognized chemical name fragments using a list of stop words to eliminate erroneous chemical name fragments.
30. (Currently Amended) ~~A system~~ The apparatus as in claim 19, where chemical name

These amendments are for discussion purposes only. /KRS/

fragments are further recognized by using common chemical word endings.

31. (Currently Amended) ~~A system~~ The apparatus as in claim 19, where application of said regular expressions and rules results in punctuation characters being one of maintained or removed between chemical name fragments as a function of context.
32. (Currently Amended) ~~A system~~ The apparatus as in claim 19, where said regular expressions comprise a plurality of patterns, individual ones of which are comprised of at least one of characters, numbers and punctuation.
33. (Currently Amended) ~~A system~~ The apparatus as in claim 32, where the punctuation comprises at least one of parenthesis, square bracket, hyphen, colon and semi-colon.
34. (Currently Amended) ~~A system~~ The apparatus as in claim 32, where the characters comprise at least one of upper case C, O, R, N and H.
35. (Currently Amended) ~~A system~~ The apparatus as in claim 32, where the characters comprise strings of at least one of lower case xy, ene, ine, yl, ane and oic.
36. (Currently Amended) ~~A system~~ An apparatus as in claim 19, further comprising an input tokenizer ~~unit~~ module configured to receive documents to be processed to provide a sequence of tokens.
37. (Currently Amended) ~~A computer program product for storing in a computer readable medium embodying~~ form a set of computer program instructions for directing at least one

These amendments are for discussion purposes only. /KRS/

computer to process text of a text document, comprising:

instructions to parse the text of the text document to recognize chemical name fragments; instructions to recognize any substructures present in the chemical name fragments;

instructions to extract information associated with the recognized chemical name fragments and substructures of the text document and index the extracted information in a text index;

instructions to determine structural connectivity information of the chemical name fragments and recognized substructures;

instructions to index representations of the chemical name fragments and the recognized substructures in association with the determined structural connectivity information into a plurality of chemical connectivity tables;

instructions to store the text index in association with the indexed representations in a searchable index; and

instructions to provide a graphical user interface to search the searchable index, where the search comprises entering one or more chemical fragment names and entering one or more substructures in a representation form, where the entering is by at least one of text form or graphical selection.

38. (Currently Amended) A computer program product readable medium embodying computer program instructions as in claim 37, wherein the instructions to extract information further extract keywords from the text document and index the keywords in the text index,

These amendments are for discussion purposes only. /KRS/

and wherein the search comprises selecting a graphical representation of one or more substructures and additionally entering at least one keyword.

39. (Currently Amended) A The computer program product readable medium embodying computer program instructions as in claim 37, wherein the instructions to extract information further extract keywords from the text document and index the keywords in the text index, and wherein the search comprises additionally entering at least one keyword and at least one of fragment connectivity and substructure connectivity.

40. (Cancelled)

41. (Cancelled)

42. (Currently Amended) A The computer program product readable medium embodying computer program instructions as in claim 37, wherein the search further comprises entering at least one search term, and where a search results in an intersection of the indexed representations and the text index, to identify at least one document that contains a reference to a corresponding chemical compound.

43. (Currently Amended) A system comprising a plurality of computers at least two of which are coupled together through a data communications network, said system comprising:

a unit tokenizer and a token processing unit configured to parse text of a text document to recognize chemical name fragments; a unit to recognize any substructures present in the chemical name fragments;

These amendments are for discussion purposes only. /KRS/

a unit the token processing unit configured to extract information associated with the recognized chemical name fragments and substructures from the text document and index the extracted information in a text index;

a unit the token processing unit configured to determine structural connectivity information of the chemical name fragments and recognized substructures;

a unit the token processing unit configured to index representations of the chemical name fragments and the recognized substructures in association with the determined structural connectivity information into a plurality of chemical connectivity tables;

a unit the token processing unit configured to store the text index in association with the indexed representations in a searchable index; and

a unit to provide searcher unit and a graphical user interface configured to search the searchable index, where the search comprises entering one or more chemical fragment names and entering one or more substructures in a representation form, where the entering is by at least one of text form or graphical selection.

44. (Currently Amended) A. The system as in claim 43, wherein the the search further comprises entering at least one search term, and wherein a search results in an intersection of a the indexed representations and the text index, to identify at least one document that contains a reference to a corresponding chemical compound.

These amendments are for discussion purposes only. /KRS/

45. (Currently Amended) A The system as in claim 43, where said token processing unit that  
determines structural connectivity information is further configured to looks up recognized  
fragments and substructures in a structure dictionary.

46. (Currently Amended) A The system as in claim 43, where the representations comprise MOL  
type representations and SMILES type representations.